Analiza si vizualizarea datelor

Nicoleta ROGOVSCHI

nicoleta.rogovschi@parisdescartes.fr

Recall:
Techniques of dimension reduction

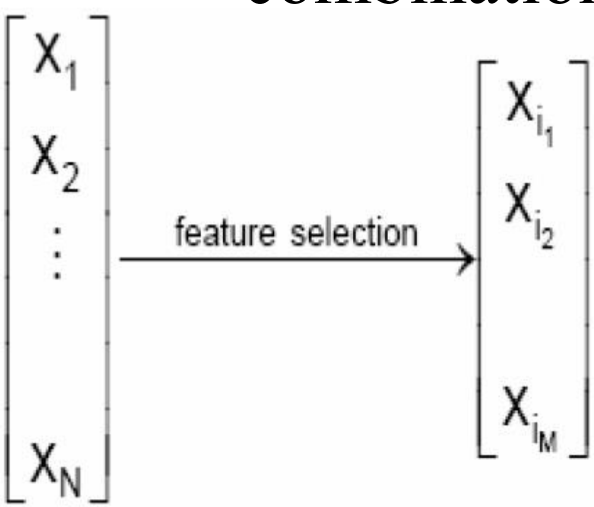
Dimension Reduction

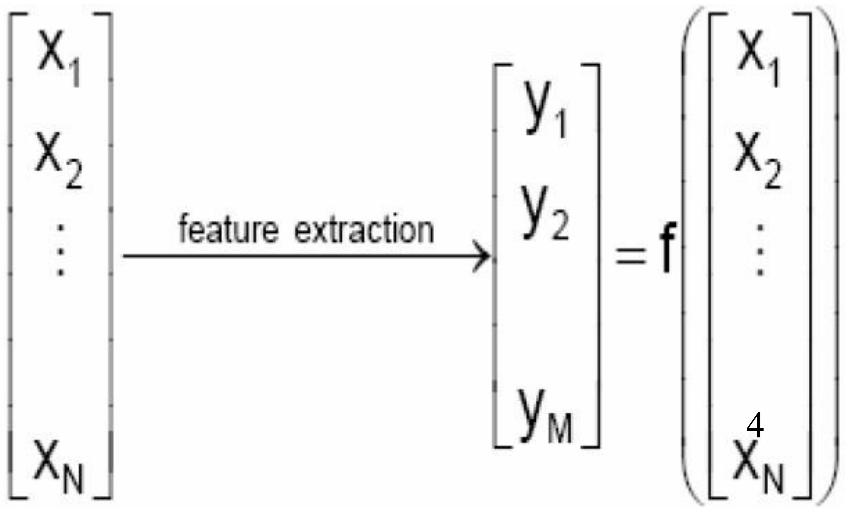
There are a variety of techniques of dimension reduction:

- Linear vs. non-linear
- Deterministic vs. probabilistic
- Supervised vs. unsupervised

Dimension Reduction

- Dimension reduction : Methodologies
  - Feature selection: choosing a subset of all the feature
  - Feature extraction (« feature extraction ») : creating a subset of new features by combinations





Dimension reduction
via
Feature selection

Feature selection

Problem : most of the evaluation criteria are not monotonous

Use of sub-optimal methods: :

- *Sequential Forward Selection (SFS)*
- *Sequential Backward Selection (SBS)*
- *Bidirectional Selection (BS)*

Dimension reduction via Feature extraction

Dimension reduction via feature extraction

Two main types of methods :

- **Linear Methods**
    - Principal Components Analysis (PCA)
    - Linear Discriminant Analysis (LDA)
    - Multi-Dimensional Scaling (MDS)
    - …
- **Non-Linear Methods**
    - Isometric feature mapping (Isomap)
    - Locally Linear Embedding (LLE)
    - Kernel PCA
    - Spectral clustering
    - Supervised methods (S-Isomap)
    - …

Linear Discriminant Analysis (LDA)

Outline

- Introduction and definitions
- ADL for 2 classes
- ADL for multiple classes
- Example
- PCA vs ADL
- Conclusions

Dimension reduction via feature extraction

Two main types of methods :

- **Linear Methods**
  - Principal Components Analysis (PCA)
  - Linear Discriminant Analysis (LDA)
  - Multi-Dimensional Scaling (MDS)
  - …
- **Non-Linear Methods**
  - Isometric feature mapping (Isomap)
  - Locally Linear Embedding (LLE)
  - Kernel PCA
  - Spectral clustering
  - Supervised methods (S-Isomap)
  - …

Introduction

- Linear discriminant analysis (LDA)
  - A method for the analysis of high-dimensional data in the case of supervised learning (classes (labels) are available in the data set)
  - Find an optimal low-dimension space such that when points are projected, the data from different classes are well separated
  - Useful for feature extraction to facilitate classification

Introduction

- LDA tries to determine the contribution of variables that explain the membership of individuals to groups.

- The linear discriminant analysis can also affect new objects to groups.

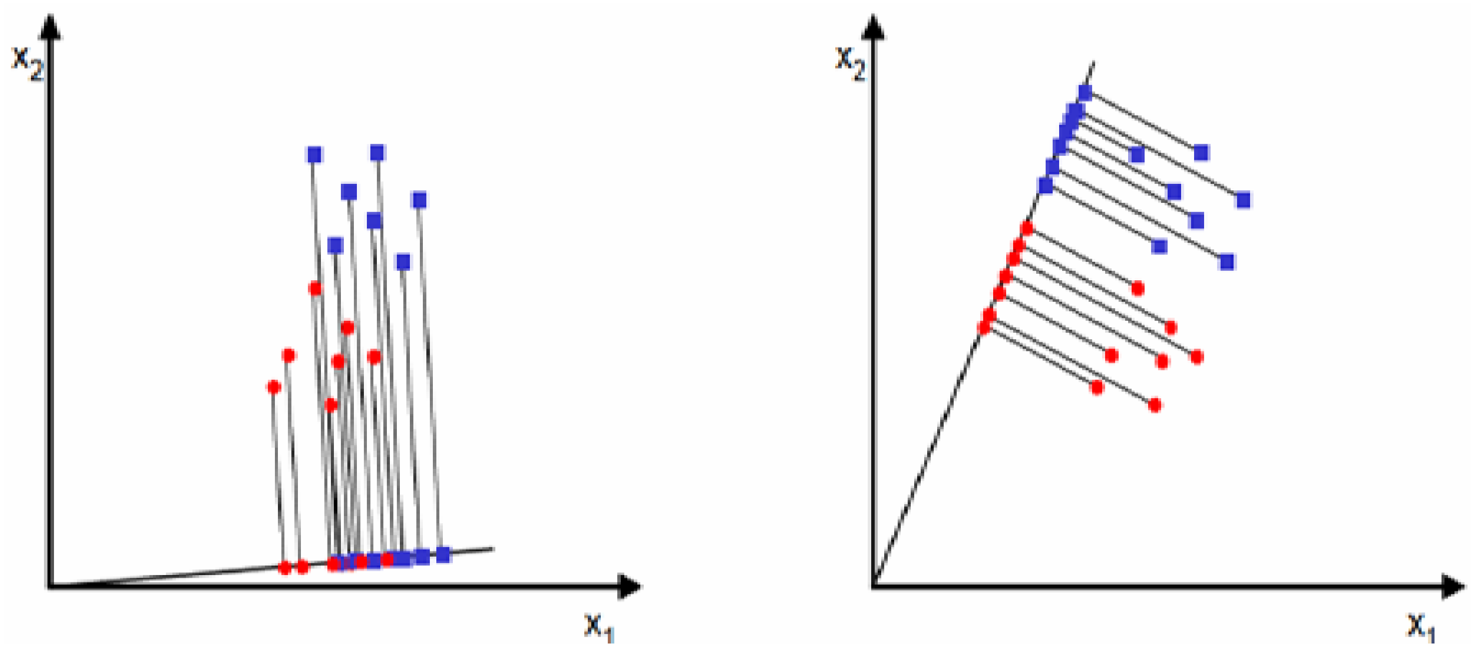
LDA for 2 classes

- The main objective of LDA is to achieve dimensionality reduction while preserving as much as possible the discriminatory information of each class
  - Assume that we have a set of D-dimensional samples $\left\{x^{(1}, x^{(2}, \ldots, x^{(N}\right\}$, $\boldsymbol{N_1}$ of which belong to class $\boldsymbol{r_1}$ and N_2 to class $\boldsymbol{r_2}$.
  - We seek to obtain a scalar **y** by projecting the samples $\boldsymbol{x}$ onto a line $\boldsymbol{w}$

$$y = w^T x$$

LDA for 2 classes

- Of all the possible lines we would like to select the one that maximizes the separability of the scalars



LDA for 2 classes

To find a good projection vector, we need to define a mesure of separation between the projections.

– The mean vector of each class in ***x*** and ***y*** feature space is:

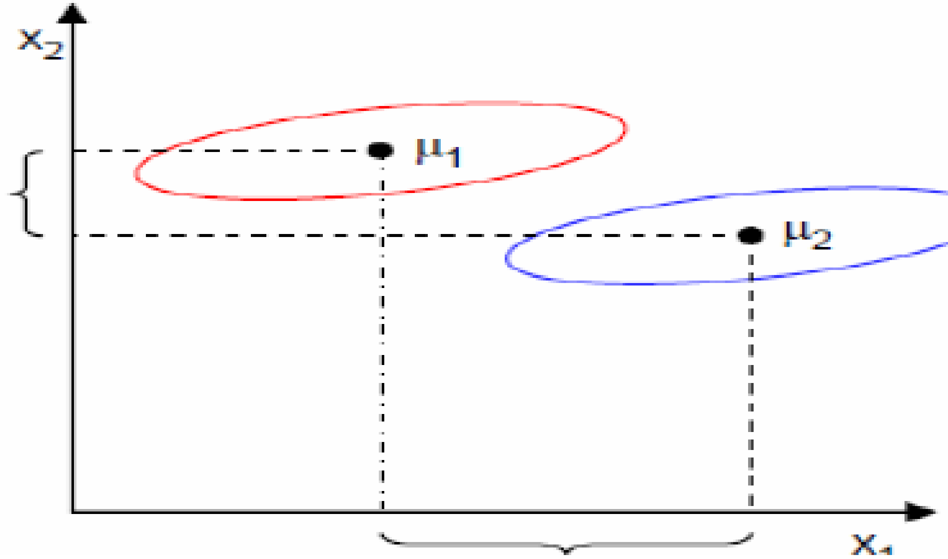
$$\mu_i = \frac{1}{N_i}\sum_{x \in r_i} x \quad \text{et} \qquad \tilde{\mu}_i = \frac{1}{N_i}\sum_{y \in r_i} y = \frac{1}{N_i}\sum_{x \in r_i} w^T x = w^T \mu_i$$

LDA for 2 classes

- We can then choose the distance between the projected means as our objective function :

$$J(w) = \left|\tilde{\mu}_1 - \tilde{\mu}_2\right| = \left|w^T(\mu_1 - \mu_2)\right|$$

- However, the distance between the projected means is not a very good measure since it does not take into account the standard deviation within the classes.



LDA for 2 classes

- The solution proposed by Fisher is to maximize a function that represents the difference between the means, normalizes by a measure of the within-classe scatter
  - For each class we define the scatter, an equivalent of the variance, as:

$$\tilde{s}_i^{\,2} = \sum_{y \in r_i} (y - \tilde{\mu}_i)^2$$

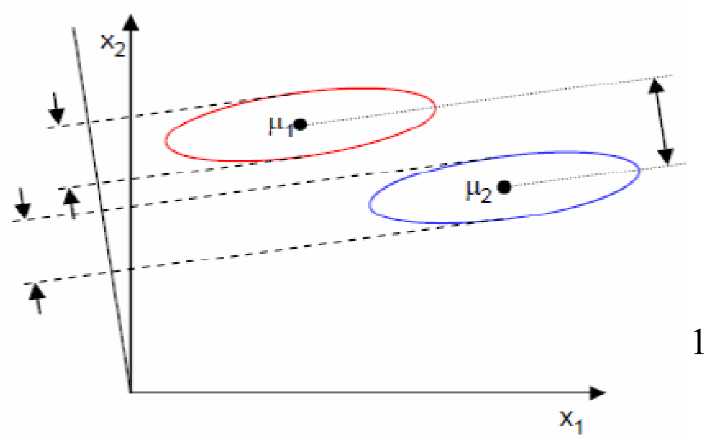
Where the quantity $(\tilde{s}_1^{\,2} + \tilde{s}_2^{\,2})$ is called the **within-class scatter** of the projected examples.

LDA for 2 classes

The Fisher linear discriminant is defined as the linear function $\boldsymbol{w}^T\boldsymbol{x}$ that maximise the criterion function :

$$J(w) = \frac{\left|\tilde{\mu}_1 - \tilde{\mu}_2\right|^2}{\tilde{s}_1^{\,2} + \tilde{s}_2^{\,2}}$$

Therefore, we will be looking for a projection where examples from the same class are projected very close to each other and, at the same time, the projected means are as farther apart as possible



LDA for 2 classes

- To find the optimum projection w*, we need to express J(w) as an explicit function of **w**
- We define a measure of the scatter in multivariate feature space **x**, which are **scatter matrices**

$$S_i = \sum_{x \in r_i} (x - \mu_i)(x - \mu_i)^T$$

$$S_w = S_1 + S_2$$

where S_W is called the **within-class scater matrix**.

LDA for 2 classes

- The scatter plot of the projection ***y*** can be expressed as a function of the scatter matrix in feature space ***x***

$$\tilde{s}_i^{\,2} = \sum_{y \in r_i} (y - \tilde{\mu}_i)^2 = \sum_{x \in r_i} (w^T x - w^T \mu_i)^2 = \sum_{x \in r_i} w^T (x - \mu_i)(x - \mu_i)^T w = w^T S_i w$$

$$w^T S_w w = \tilde{s}_1^{\,2} + \tilde{s}_2^{\,2}$$

- In a similar way, the difference between the projected means can be expressed in terms of the means in the original feature space

$$(\tilde{\mu}_1 - \tilde{\mu}_2)^2 = (w^T \mu_1 - w^T \mu_2)^2 = w^T \underbrace{(\mu_1 - \mu_2)(\mu_1 - \mu_2)}_{S_B} w = w^T S_B w$$

LDA for 2 classes

- The matrix S_B is called the between-class scatter.

- We can express the Fisher criterion in terms of S_W and S_B as :

$$J(w) = \frac{w^T S^B w}{w^T S_w w}$$

ADL pour 2 classes

- To find the maximum of J(W) we derive and equate to zero

$$\frac{d}{dw}\left[J(w)\right] = \frac{d}{dw}\left[\frac{w^T S_B w}{w^T S_W w}\right] = 0 \Rightarrow$$
$$\Rightarrow \left[w^T S_W w\right]\frac{d\left[w^T S_B w\right]}{dw} - \left[w^T S_B w\right]\frac{d\left[w^T S_W w\right]}{dw} = 0 \Rightarrow$$
$$\Rightarrow \left[w^T S_W w\right]2S_B w - \left[w^T S_B w\right]2S_W w = 0$$

- We divide by $w^T S_W w$

$$\left[\frac{w^T S_W w}{w^T S_W w}\right]S_B w - \left[\frac{w^T S_B w}{w^T S_W w}\right]S_W w = 0 \Rightarrow$$
$$\Rightarrow S_B w - J S_W w = 0 \Rightarrow$$
$$\Rightarrow S_W^{-1} S_B w - J w = 0$$

LDA for 2 classes

- Solving the generalized eigenvalue problem $S_W^{-1} S_B w = J w$ yields to

$$w^* = \arg\max_w \left\{ \frac{w^T S_B w}{w^T S_w w} \right\} = S_w^{-1} (\mu_1 - \mu_2)$$

- Which represents the Fisher's Linear Discriminant (1936).

LDA for Multiple Classes

- We can easily generalize Fisher's LDA for C-classes problem.
  - Instead of one projection **y**, we will seek now (C-1) projections $[y_1, y_2, \ldots, y_{C-1}]$. We will have (C-1) projection vectors w_i which can be arranged by columns into a projection matrix $W=[w_1|w_2|\ldots|w_{C-1}]$:

$$y_i = w_i^T x \Rightarrow y = W^T x$$

LDA for Multiple Classes

- Generalizations :
  - The generalization of the within-class scater is

$$S_W = \sum_{i=1}^{c} S_i$$

where $S_i = \sum_{x \in r_i} (x - \mu_i)(x - \mu_i)^T$ and $\mu_i = \frac{1}{N_i} \sum_{x \in r_i} x$

  - The generalization of the between-class scater is

$$S_B = \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T$$

where $\mu = \frac{1}{N} \sum_{\forall x} x = \frac{1}{N} \sum_{x \in r_i} N_i \mu_i$

$S_T = S_B + S_W$ is called the **total scatter matrix.**

LDA for Multiple Classes

- Similarly, we can define the mean vector and scatter matrices for the projected samples as :

$$\tilde{\mu}_i = \frac{1}{N_i} \sum_{y \in r_i} y \qquad \tilde{S}_W = \sum_{i=1}^{C} \sum_{y \in r_i} (y - \tilde{\mu}_i)(y - \tilde{\mu}_i)^T$$

$$\tilde{\mu} = \frac{1}{N} \sum_{\forall y} y \qquad \tilde{S}_B = \sum_{i=1}^{C} N_i (\tilde{\mu}_i - \mu)(\tilde{\mu}_i - \mu)^T$$

- From our derivation for the 2-class problem, we can write:

$$\tilde{S}_W = W^T S_W W$$

$$\tilde{S}_B = W^T S_B W$$

LDA for Multiple Classes

- Recall that we are looking for a projection that maximizes the ratio of between-class to within-class scatter. Since the projection is no longer a scalar (it has C-1 dimensions), we then use the determinant of the scatter matrices to obtain a scalar objective function:

$$J(W) = \frac{|\tilde{S}_B|}{|\tilde{S}_W|} = \frac{|W^T S_B W|}{|W^T S_W W|}$$

- And we will seek the projection matrix W* that maximizes this ratio

ADL multi-classes

- It can be shown that the optimal projection matrix W* is the one whose columns are the eigenvectors corresponding to the largest eigenvalues of the following generalized eigenvalue problem :

$$W^* = \left[w_1^* \middle| w_2^* \middle| \ldots \middle| w_{C-1}^* \right] = \arg\max_W \left\{ \frac{\left| W^T S_B W \right|}{\left| W^T S_w W \right|} \right\} = (S_B - \lambda_i S_W) w_i^* = 0$$

Example

- We have a two-dimensional dataset

  $X1=(x_1, x_2)=\{(4,1),(2,4),(2,3),(3,6),(4,4)\}$

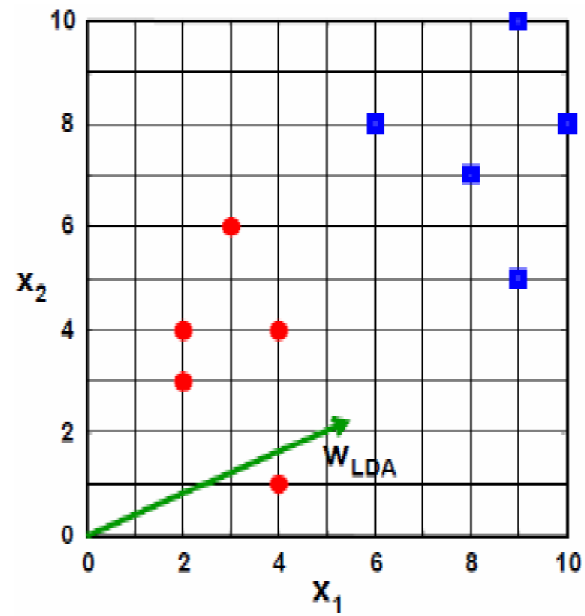
  $X2=(x_1, x_2)=\{(9,10),(6,8),(9,5),(8,7),(10,8)\}$

- Solution:
  - The class statistics are:

$$S_1 = \begin{bmatrix} 0.80 & -0.40 \\ -0.40 & 2.60 \end{bmatrix}; \qquad S_2 = \begin{bmatrix} 1.84 & -0.04 \\ -0.04 & 2.64 \end{bmatrix};$$

$$\mu_1 = \begin{bmatrix} 3.00 & 3.60 \end{bmatrix}; \qquad \mu_2 = \begin{bmatrix} 8.40 & 7.60 \end{bmatrix};$$

  - The within and between classes scatter are :



$$S_B = \begin{bmatrix} 29.16 & 21.60 \\ 21.60 & 16.00 \end{bmatrix}; \qquad S_W = \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{bmatrix};$$

Example

– The LDA projection is obtained as the solution of the generalized eigenvalue problem :

$$S_W^{-1} S_B v = \lambda v \Rightarrow \left| S_W^{-1} S_B - \lambda I \right| = 0 \Rightarrow \begin{vmatrix} 11.89 - \lambda & 8.81 \\ 5.08 & 3.76 - \lambda \end{vmatrix} = 0 \Rightarrow \lambda = 15.65$$

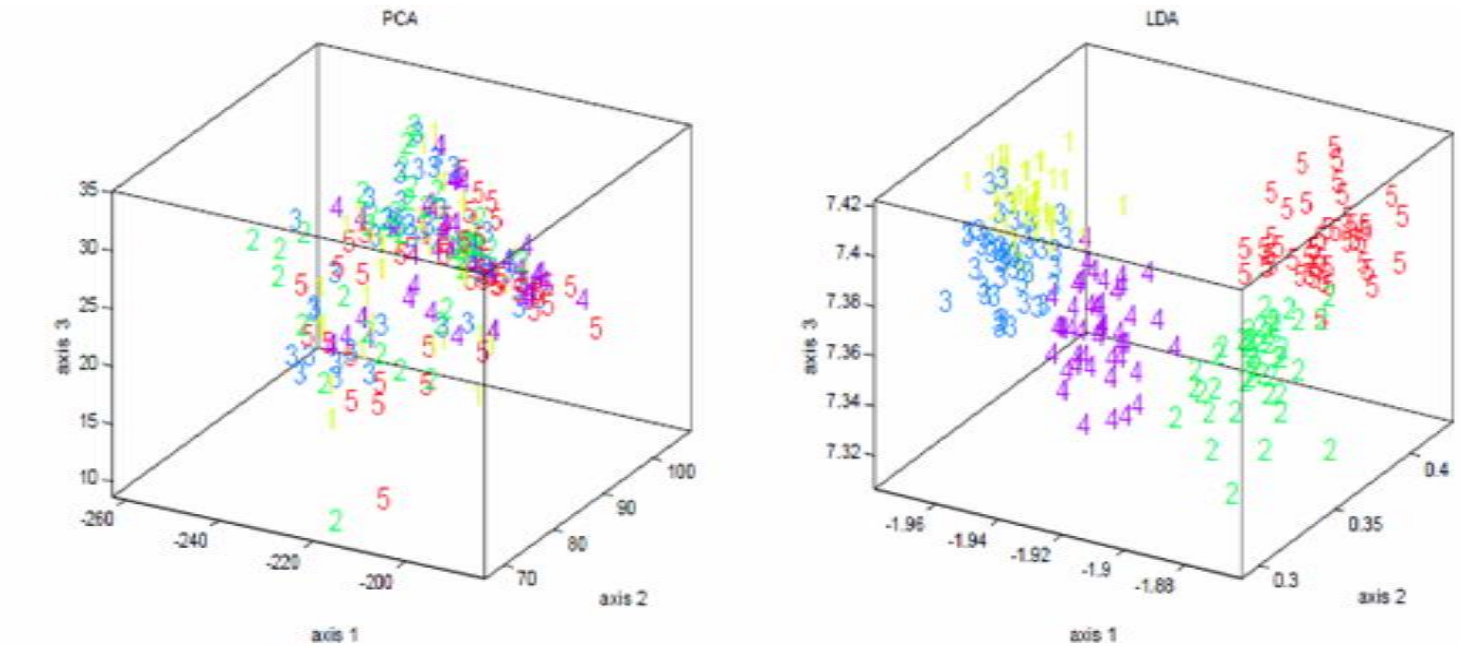
$$\begin{bmatrix} 11.89 & 8.81 \\ 5.08 & 3.76 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 15.65 \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \Rightarrow \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0.91 \\ 0.39 \end{bmatrix}$$

– Or directly by :

$$W^* = S_W^{-1}(\mu_1 - \mu_2) = \begin{bmatrix} -0.91 & -0.39 \end{bmatrix}^T$$

PCA vs LDA

- We have a dataset of 5 types of coffee beans (60-dimensional feature vector)



PCA vs LDA

- For small data sets PCA gives better results than LDA.

- When the number of examples is large enough and representative for each class, the ADL shows better performance than the PCA.

Conclusions

- ADL is a simply well known method for mining high dimensional data, when the class labels are available.

- It is a linear method for dimensionality reduction by projecting the original data to C-1 dimensional space

Conclusions

- There are a number of limitations in the classical LDA
- There are many extensions of classical LDA (generalized LDA, non-parametric LDA, orthonormal LDA) which tend to overcome these limitations.

Multivariate analyses

- The multidimensional or multivariate analysis methods allow to process simultaneously many variables that characterize individuals of the study.

These are mainly descriptive purposes tools:

- designed to obtain a synthetic representation of a data table
- extracting the maximum of "information", more exactly variability, variance or inertia
- in consideration with the minimum of distortion compared to the original data

The aim of multivariate analysis methods

- When we are in the presence of 2 or 3 variables, it is possible to have a graphical representation that restores all the information. This is no longer true if one is interested in more than 3 variables.

- The principle of data analysis is to plot on a 2 dimensional graph (thanks to a projection) all the observations.

- However, the selected axes do not correspond to either of the variables but are virtual lines, obtained from combinations of the variables and calculated to get as close to all points. Each point is projected on the plan.

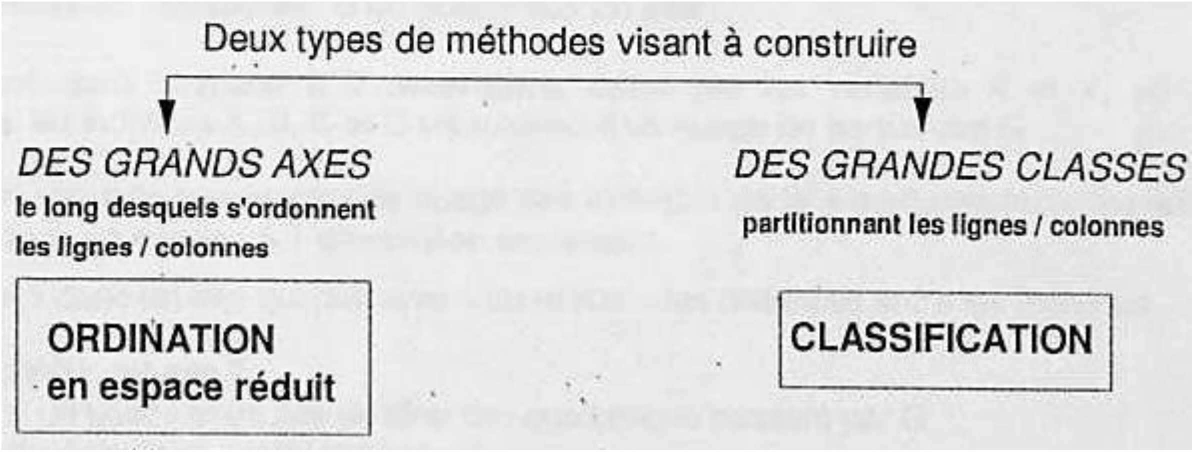
- The choice of axes is made so that the graph summarizes the data, minimizing the loss of information.

Types of multivariate analysis methods

In general there are two major types of methods:

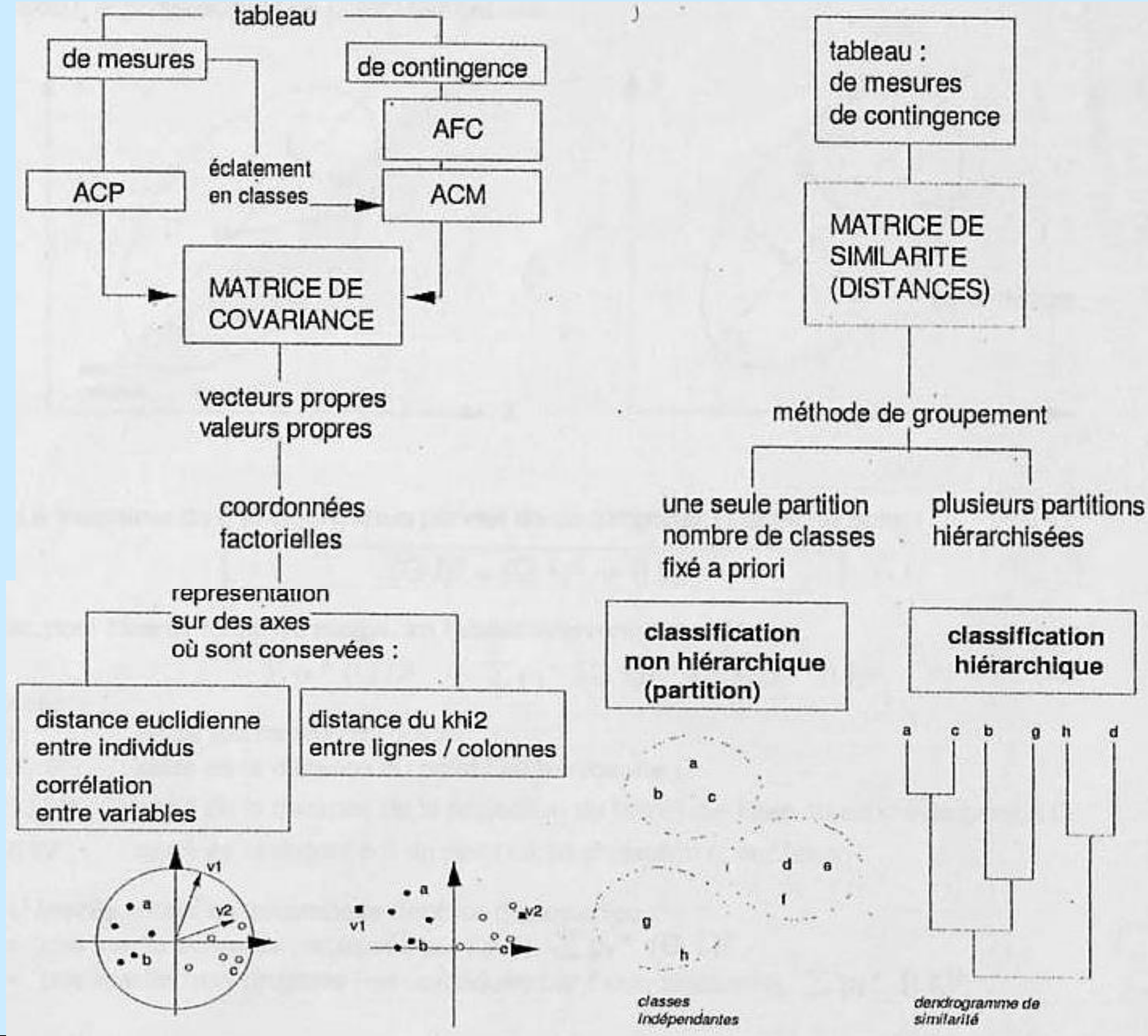
- Ordination (factorial methods) : are used to identify major axes along which the objects and / or variables are arranged

- Classification (hierarchical or not) : define large classes in which objects fall is (rarely variables) according to similarity criteria



ORDINATION
in small space

CLASSIFICATION



The data analysis methods

There are three factorial data analysis methods:

- PCA : The Principal Component Analysis: dedicated for quantitative variables.

- CFA : The **correspondence factor analysis** is applied to two categorical variables (nominal).

- MCFA : **Multiple correspondence factoral analysis** generalizes CFA to any number of variables.

The objectives of the PCA

- What are the variables that are positively related to each other?

- Which ones are opposed?

- Compared to observations, we try to assess their similarity and dissimilarity, and to highlight homogeneous groups of observations.

Principles of the PCA

- Was developed in 1933 by H.Hotelling
- The oldest methods of data analysis
- Principle : obtain an approximate representation of the cloud of the *n* individuals in a subspace of low dimension. Summarize at best a data table represented by a matrix X with *n* lines and *p* columns.

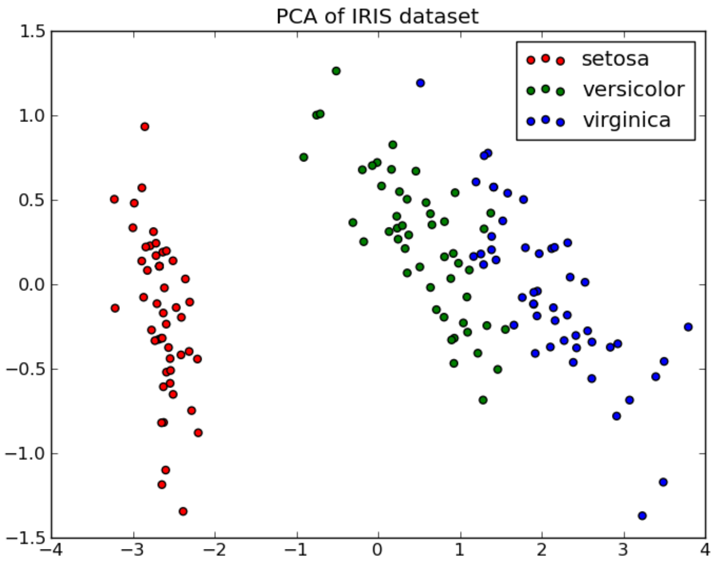
$$R^p \to R^2$$

Example : Iris of Fisher

- Number of observations: 150 (50 in all 3 classes)

- Number of variables: 4 numerical predictive variables and the class

- Descriptions of variables:
  1. the sepals length in cm
  2. the sepals width in cm
  3. the petals length in cm
  4. the petals width in cm
  5. the classes (the labels):
     -- Iris Setosa
     -- Iris Versicolour
     -- Iris Virginica

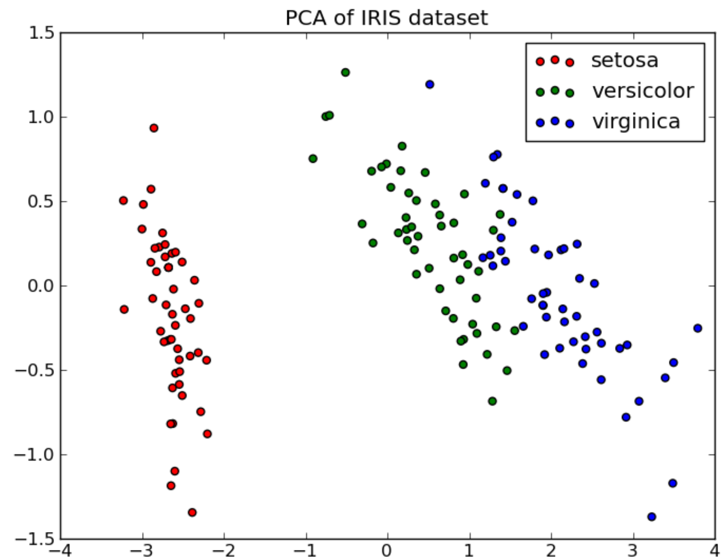


Principal Component Analysis (PCA)

Iris of Fisher

5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
…
6.0,2.2,4.0,1.0,Iris-versicolor
6.1,2.9,4.7,1.4,Iris-versicolor
5.6,2.9,3.6,1.3,Iris-versicolor
6.7,3.1,4.4,1.4,Iris-versicolor
5.6,3.0,4.5,1.5,Iris-versicolor
5.8,2.7,4.1,1.0,Iris-versicolor
6.2,2.2,4.5,1.5,Iris-versicolor
…
6.0,3.0,4.8,1.8,Iris-virginica
6.9,3.1,5.4,2.1,Iris-virginica
6.7,3.1,5.6,2.4,Iris-virginica
6.9,3.1,5.1,2.3,Iris-virginica
5.8,2.7,5.1,1.9,Iris-virginica
6.8,3.2,5.9,2.3,Iris-virginica
6.7,3.3,5.7,2.5,Iris-virginica
6.7,3.0,5.2,2.3,Iris-virginica
6.3,2.5,5.0,1.9,Iris-virginica
6.5,3.0,5.2,2.0,Iris-virginica
6.2,3.4,5.4,2.3,Iris-virginica
5.9,3.0,5.1,1.8,Iris-virginica

PCA



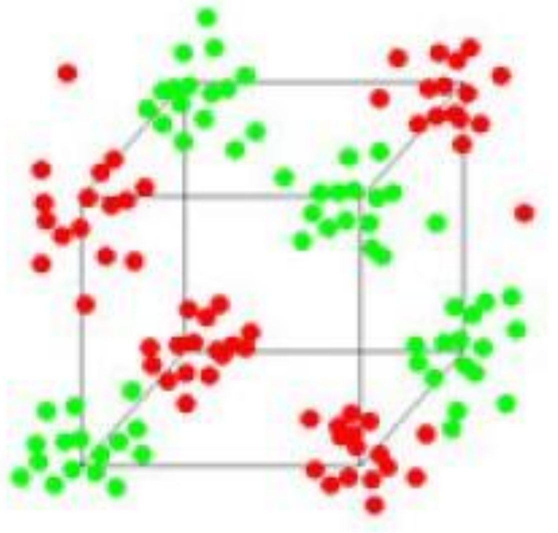
$R^4 \longrightarrow R^2$

Principal Component Analysis (PCA)

«Madelon» dataset

4400 observations x 500 variables

PCA



$R^{500} \longrightarrow R^3$

PCA

Let $X = \{x_1, x_2, \ldots, x_n\}$ of R^p the set of observations (population of *n* individuals with *p* features).

$$X = (x^1, \ldots, x^p) = \begin{bmatrix} x_1^1 & x_1^2 & & \cdots & x_1^p \\ x_2^1 & x_2^2 & & & \\ & & \ddots & & \\ \vdots & & & x_i^j & \vdots \\ & & & \ddots & \\ x_n^1 & & & \cdots & x_n^p \end{bmatrix}$$

Variable
A column of the table

$$x^j = \begin{pmatrix} x_1^j \\ x_2^j \\ \cdots \\ \cdots \\ \cdots \\ x_N^j \end{pmatrix}$$

Individual
A row table

$$x_i = \left(x_i^1, x_i^2, \ldots\ldots, x_i^p\right)$$

$$\bar{x} = \frac{1}{n} \begin{pmatrix} \sum_{k=1}^{n} x_k^1 \\ \sum_{k=1}^{n} x_k^2 \\ \cdots \\ \cdots \\ \cdots \\ \sum_{k=1}^{n} x_k^p \end{pmatrix}$$

Mean

The dispersion of the values of a variable around its mean is measured by its variance :

$$\operatorname{var}(x^j) = \frac{1}{n} \sum_{k=1}^{n} \left(x_k^j - \bar{x}^j\right)^2$$

PCA

To study the mutual influence between two variables x^i and x^j we introduce the covariance :

$$\operatorname{cov}(x^i, x^j) = \frac{1}{n}\sum_{k=1}^{n}(x_k^i - \bar{x}^i)(x_k^j - \bar{x}^j)$$

The correlation of x^i and x^j is defined by :

$$r(x^i, x^j) = \frac{\operatorname{cov}(x^i, x^j)}{s^i s^j}$$

PCA

There exists two types of PCA :

- PCA non normalized (centered)

  - Data centering

$$y = x - \bar{x}$$

- PCA normalized

  - Centering and data reduction

$$y = \frac{x - \bar{x}}{s}$$

PCA non normalized

If the input data are heterogeneous compared to their average, the matrix $X = \{x_1, x_2, \ldots, x_n\}$ is transformed :

$$X = \{x_1, x_2, \ldots, x_n\} \rightarrow Y = \{x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x}\}$$

The covariance matrix is defined by : $M_{\text{cov}} = \frac{1}{n}\left(Y^t Y\right)$

$$M_{\text{cov}} = \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1n} \\ s_{21} & s_2^2 & & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_n^2 \end{bmatrix}$$

where s_{ij} is the covariance of x_i and x_j

Diagonalization

$$M_{\text{cov}} = \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1n} \\ s_{21} & s_2^2 & & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_n^2 \end{bmatrix} \xrightarrow[\textit{diagonalization}]{M_{\text{cov}} = U\lambda U^{-1}}$$

$$\lambda_{p\times p} = \begin{bmatrix} \lambda_1 & 0 & \ldots & \ldots & 0 \\ 0 & \lambda_2 & \ldots & \ldots & 0 \\ : & : & \lambda_i & : & : \\ : & : & : & : & : \\ 0 & \ldots & \ldots & \ldots & \lambda_p \end{bmatrix}$$

$$U_{p\times p} = \begin{bmatrix} u_{1,1} & u_{1,2} & \ldots & \ldots & u_{1,p} \\ u_{2,1} & u_{2,2} & \ldots & \ldots & u_{2,p} \\ : & : & u_{i,j} & : & : \\ : & : & : & : & : \\ u_{p,1} & \ldots & \ldots & \ldots & u_{p,p} \end{bmatrix}$$

Main axes

We call principal axes of inertia, the *p* eigenvectors M_{cov} .

The first main axis is the eigenvector u_1 corresponding to the largest eigenvalue λ_1 of M_{cov}.

The inertia explained by this axis is λ_1 .

The sub-space of *p* dimensions which explains the greater inertia contains the *p* eigenvectors $(u_1, u_2, \ldots, u_p)$ of M_{cov}.

The inertia explained by this sub-space is equal to : $\sum_{i=1}^{p} \lambda_i$

Principal components

Calculation of coordinates of points on main axes :

The principal components are obtained by :

$$c_k = F_k(x_i) = \langle u_k, x_i \rangle = \sum_{j=1}^{p} u_{k,j} x_i^j$$

k^th principal component

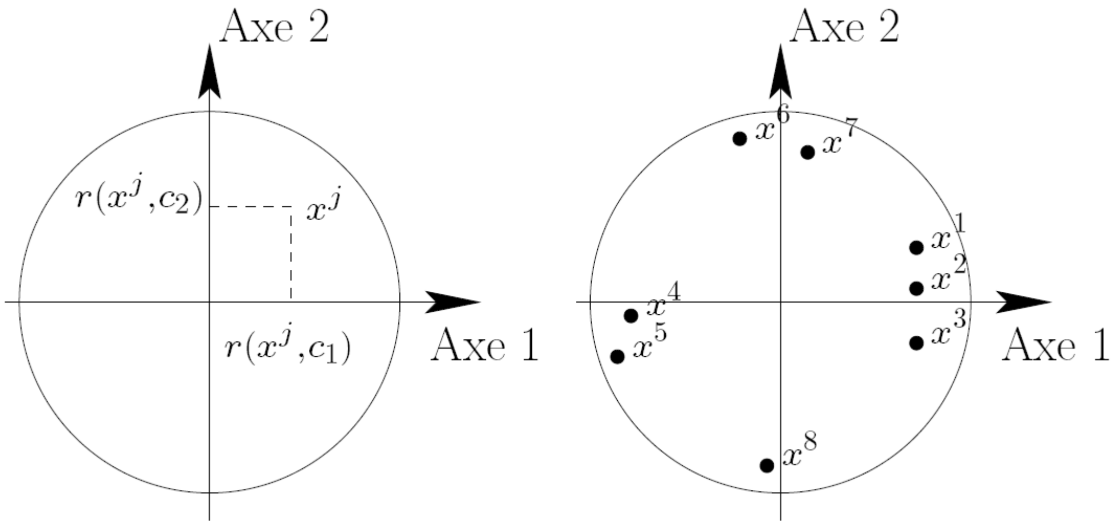
The principal components can be considered as new variables, linear combinations of the original variables, uncorrelated and of maximum variance.

Correlation circle

To interpret the relationship between the original variables and the factorial axes, we define the following correlation coefficient :

$$r(c_k, x^j) = \frac{\frac{1}{n}\sum_{i=1}^{n} F_k(x_i)\left(x_i^j - \bar{x}_j\right)}{s_j\sqrt{\frac{\lambda_k}{n}}}$$

$r(u_k,x^j)$ is none other than the correlation coefficient between $F_k(x_i)$ of λ_k inertia and λ_k /n variance, and the x_i^j of variance σ_j.



The first main component is positively correlated with the variables 1, 2 and 3, anticorrelated with 4 and 5 variables and uncorrelated with 6,7 and 8.

We must interpret the proximity of the points if they are close to the circumference.

PCA normalized

PCA normalized

We assume that the original variables, are not only heterogeneous compared to their average, but also compared to their dispersion and the type (measurement units).

So we transform each variable to a common framework of comparability: the variables must be of variance one and zero mean.

We transform the matrix $X = \{x_1, x_2, \ldots, x_n\}$

$$X = \{x_1, x_2, \ldots, x_n\} \rightarrow Y = \left\{ \frac{x_1 - \bar{x}}{s}, \frac{x_2 - \bar{x}}{s}, \ldots, \frac{x_n - \bar{x}}{s} \right\}$$

The data are centered and reduced.

PCA normalized

We define the correlation matrix by :

$$M_{corr} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1n} \\ r_{21} & 1 & & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \dots & 1 \end{bmatrix}$$

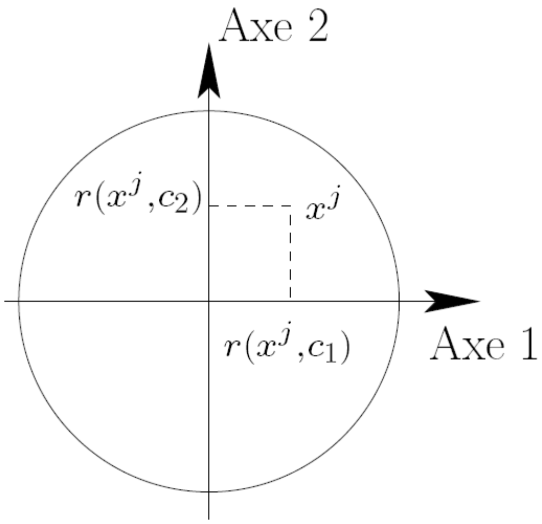
where r_{xy} is the linear correlation coefficient of the variables X and Y.

The normalized PCA consists in diagonalizing M_{corr} instead of M_{cov}.

PCA normalized

The interpretation of the axes is done by the study of the correlations between the principal component defining this axis and the variables of the original data table:

$$r(c^k, x^j) = \sqrt{\lambda_k}\, u_j^k$$

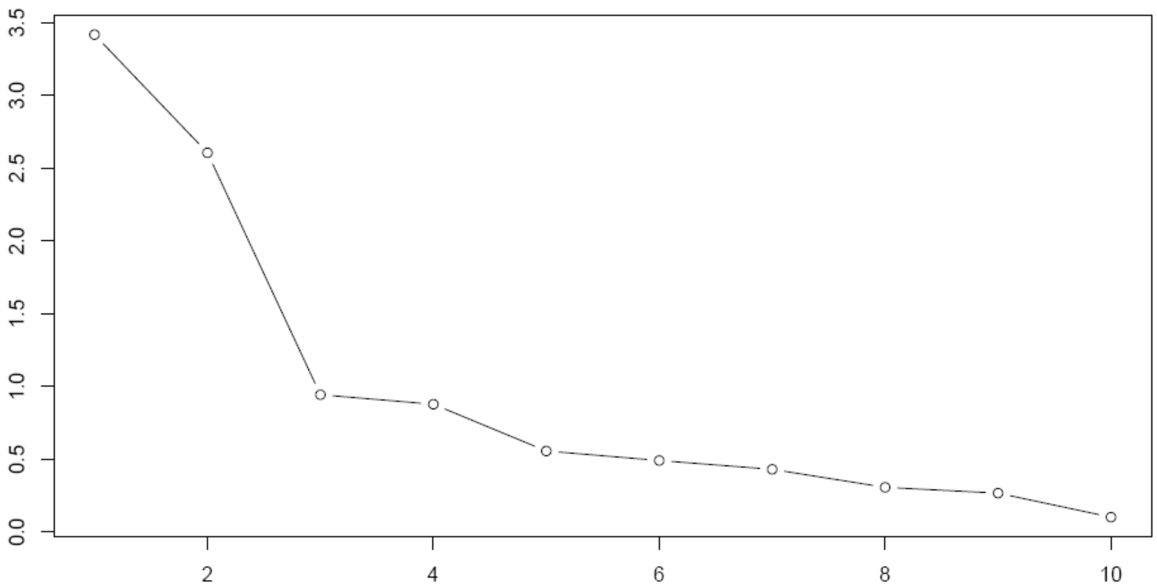


Number of axes to remember

- **Kaiser criterion (reduced centered variables)**

One retains only the axes associated to the eigenvalues upper than 1, i.e to whose variance is greater than that of the original variables.

- We seek a bend in the graph of the eigenvalues



PCA

Steps of computing:

- Centering the data matrix
- Diagonalization of the obtained matrix
- Calculation of the principal components
- Graphic representation

Size factor

- Variables must be all on the same side of one factor axis. Such an arrangement appears when all variables are positively correlated with each other. If for an individual, a variable takes a high value, all other variables are also taking a strong value.

- This characteristic is present most often on the first factorial axis. We speak of the effect of "size" or size factor.